

Случайные процессы. Прикладной поток.

Практическое задание 5.

Правила:

- Выполненную работу нужно отправить на почту `probability.diht@yandex.ru`, указав тему письма "[СП17] Фамилия Имя - Задание 5". Квадратные скобки обязательны. Вместо Фамилия Имя нужно подставить свои фамилию и имя.
- Прислать нужно ноутбук и его pdf-версию. Названия файлов должны быть такими: `5.N.ipynb` и `5.N.pdf`, где N — ваш номер из таблицы с оценками.
- Никакой код из данного задания при проверке запускаться не будет.

Варианты:

0. <http://ita.ee.lbl.gov/html/contrib/ClarkNet-HTTP.html>
1. <http://ita.ee.lbl.gov/html/contrib/NASA-HTTP.html>
2. <http://ita.ee.lbl.gov/html/contrib/Calgary-HTTP.html>
3. <http://ita.ee.lbl.gov/html/contrib/Sask-HTTP.html>

Нужно выбрать вариант $N \bmod 4$, где N — ваш номер из таблицы с оценками.

Задание:

1. Загрузите в Питоне данные, соответствующие вашему варианту. Возможно, будет удобнее воспользоваться библиотекой `pandas`, в частности функцией `read_csv`. Некоторые записи в данных могут иметь неправильный формат, поэтому сначала файл с данными придется преобразовать. В процессе преобразования можно проигнорировать не более 1% строк, в которых указана дата (если будете удалять, напечатайте долю удаленных строк). Строки, в которых дата не указана, нужно удалить.
2. Переведите дату в timestamp-число с помощью библиотеки `datetime`. Например, дата `24/Oct/1994:13:41:41` преобразуется в число `782995301` с помощью кода

```
from datetime import datetime

datetime.strptime("24/Oct/1994:13:41:41", "%d/%b/%Y:%H:%M:%S").timestamp()
```

Подробнее читайте в документациях к библиотекам `datetime` и `time`.
3. Запросы, происходящие в некоторые моменты времени, образуют пуассоновский процесс, смысл которого — количество запросов, обработанных сервером к текущему моменту времени. Нарисуйте графики траектории этого пуассоновского процесса за все время и за первые сутки. Нарисуйте так же аналогичные графики составного пуассоновского процесса, взяв в качестве значений количество байт (`bytes in the reply`). При построении графиков используйте код с семинара.

4. Проверьте полученный пуассоновский процесс на однородность за все время, а также за некоторые другие интервалы времени (месяц, неделя, сутки, час). Имеются ли интервалы, на которых процесс однороден? Чтобы проверить корректность работы реализованных критериев, можно проверить их на сгенерированных данных. Код для генерации можно взять с семинара.
5. Постройте оценку функции интенсивности следующим образом. Разбейте весь временной интервал на небольшие равные по длине интервалы с достаточным количеством событий в каждом интервале (например, интервала длиной в час). На каждом интервале оцените значение интенсивности процесса, считая его однородным на данном интервале. В какие моменты времени наблюдается наибольшая интенсивность?
6. Для получения дополнительных баллов исследуйте, есть ли зависимость интенсивности от типа хоста (`host`); есть ли зависимость объема запроса в байтах от времени; какой пропускной способности должны быть серверы, чтобы обрабатывать все поступающие запросы.

При проведении исследования учитывайте все особенности проверки статистических гипотез. Не забывайте делать выводы.